Ch. 26 Day 2

* Complete Warm Up

* Get group AIRFARES

* Have out books

H_0: The distribution of sales is the same for all 5 locations.

H_A: The distribution of sales is not the same for all 5 locations.

Conditions:

1. Counted Data: The data is counts of sales in different locations.
2. Randomization: Assume Representative
3. Expected Cell Frequency:

Location	1	2	3	4	5
Expected	41.2	41.2	41.2	41.2	41.2

All expected cell counts are greater than 5.

All conditions have been met for a Chi-Square goodness-of-fit test.

Mechanics:

df = $n - 1 = 4$ $\alpha = 0.05$ $\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp}$

$$\chi^2_4 = \frac{(43-41.2)^2}{41.2} + \ldots + \frac{(48-41.2)^2}{41.2} = 8.903$$

$$P - Value = P\left(\chi^2_4 > 8.903\right) = 0.0636$$

Since the P-Value is greater than alpha (0.0636 > 0.05) we fail to reject the null hypothesis.

There is not enough statistically significant evidence that the distribution of sales is different between the 5 locations

Chi-Square test #2: Homogeneity/ Independence

* Uses 2-way tables (2 variables)
* Compares 2 variables that are measured from the same sample
* Two type of tests: Homogeneity/Independence

Example: Gender vs. college location

	In-State	Out-of-State
MALE	25	36
FEMALE	37	41

Want to see if there is an association/relationship between the 2 variables

2 examples... not in notes:

p. 646 #28

Are there differences in the national origins of cars driven by students and staff?

		Driver	
		Student	Staff
Origin	American	107	105
	European	33	12
	Asian	55	47

p. 646 #29

Is there evidence of an association between gender and political affilitation?

		Political Affiliation		
		Democrat	Republican	Independent
Gender	Male	36	45	24
	Female	48	33	16

Homogeneity: Compares the distribution of several groups for the same categorical variable.

Is the distribution of grades the same throughout the years?

Is the distribution of gender the same throughout the rank in the military?

Are there differences in the distributions of type of car between students and staff?

Is the distribution of political party different between gender?

PAGE 646 #28

Independence: Examines counts from a single group to test for an association between two categorical variables

Look for: association, affect, independence, relationship

Examples:
Is there an association between grades and years?

Does gender affect someone's rank in the military?

Are type of car and type of driver independent?

Is there a relationship between political party and gender?

PAGE 646 #29

Hypotheses: Depends on the type of test.
Write the hypotheses so that they match the type of test

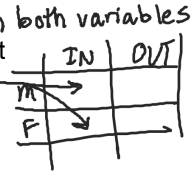
Ho: equal = no difference = independent = no association

Ha: not equal = there IS a difference = dependent = assoc.

After the Hypotheses, the rest of the test is the same for both

Conditions: Same 3 as the GOF test
1- Categorical data
2- SRS
3- All expected cell counts ≥ 5

mention both variables



conditions met --> χ^2 distribution -->
χ^2 test for Homogeneity/Independence

Mechanics:
Test statistic: same as before

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = \_\_\_\_\_ + \_\_\_\_\_ + \ldots =$$

Expected cell count = $\frac{(\text{row total})(\text{column total})}{n}$

Example: Expected counts for gender vs. college location:

Observed

	In-State	Out-of-State	
MALE	25	36	61
FEMALE	37	41	78
	62	77	139

Expected

	In-State	Out-of-State	
MALE	$\frac{61 \times 62}{139} = 27.2$	$\frac{61 \times 77}{139} = 33.8$	61
FEMALE	34.8	43.2	78
			139

P-value:

$P(\chi^2 > \text{test statistic}) = \chi^2\text{cdf}(\text{test stat}, E99, df)$

df = (r - 1)(c - 1) (r: # rows, c: # columns)

Conclusion
same 2 sentences.
* reject/fail to reject Ho....
* Sufficient/insufficient evidence that (re-copy Ha)

Calculator
- put observed in Matrix A
- Run χ^2 test (gives test stat & p-value and df)
- expected values are now in Matrix B

Example: A local college admissions officer wants to know if student choices in majors have changed through the years. He selects a random sample of freshman from 1995, 2000, and 2005 and check's their choice of major.

	1995	2000	2005
Buisness	45	79	63
Education	25	34	43
Humanities/Arts	37	46	23
Science	26	62	72
Undecided	37	29	49

Test to see whether the majors are distributed the same each year.

Hypotheses:
Ho: The choice of majors have the same distribution throughout the years.

Ha: The choice of majors do not have the same distribution throughout the years.

Conditions:
1) Categorial data
2) SRS
3) all expected cell counts ≥ 5

1) majors & the 3 years are categorical
2) Stated random
3) See table. all exp. cell counts ≥ 5

conditions met --> χ^2 distribution --> χ^2 test for Homogeneity

Test Statistic:

$\chi^2 = \sum \frac{(obs-exp)^2}{exp} = \frac{(45-47.448)^2}{47.448} + \frac{(79-69.776)^2}{69.776} + \dots$

$= 30.222$ $\quad df = 8$

P-Value: $P(\chi^2 > 30.222) = 1.931 \times 10^{-4}$

Conclusion: - We reject Ho...
- We have suff. evid. that (re copy Ha)

Example: Smoking vs. Socio-Economic Status (SES). Below is a table of a random sample of 356 PA residents, giving the Smoking status versus Socio-Economic status (SES). Test whether there is an association between the two variables. Does there appear to be a relationship between smoking and socioeconomic status?

	SES		
Smoking	High	Middle	Low
Current	51	22	43
Former	92	21	28
Never	68	9	22

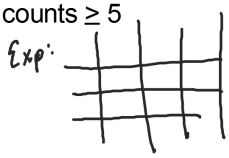
Hypotheses:
Ho: There is no association between smoking and socioeconomic status (independent) (no relationship)

Ha: There is an association between smoking and socioeconomic status (dependent) (relationship)

Conditions:
1) Categorical data
2) SRS
3) all expected counts ≥ 5

1) smoking & SES are categ.
2) stated
3) see expected table above. All exp counts ≥ 5

conditions met --> χ^2 distribution --> χ^2 test for Independence



Test Statistic:

$\chi^2 = \Sigma \frac{(obs - exp)^2}{exp} = \frac{(51 - 68.753)^2}{68.753} + \frac{(22 - 16.944)^2}{16.944} + \dots =$

$= 18.51$

P-Value:

$P(\chi^2 > 18.51 | df = 4) = 9.808 \times 10^{-4}$

Conclusion:

We reject Ho b/c p-value of 9.808×10^{-4} < alpha = 0.05.

We have sufficient evidence that there is a relationship between smoking and SES.

Worksheet in notes #1:

1) The manager of an assembly process wants to determine whether the number of defective products manufactured depends on the day of the week the articles are produced. Using the data below, is there sufficient evidence to determine if the distribution of defective products is the same throughout the work week?

Day	Mon.	Tue.	Wed.	Thur.	Fri.
Nondef.	85	90	95	95	90
Defective	15	10	5	5	10

Hypotheses:

Ho: The distribution of defective products is the same throughout the days of the week

Ha: The distribution of defective products is NOT the same throughout the days of the week

Conditions:

1) Categorical data — 1) defectiveness and days of week are categ.
2) SRS — 2) assumed representative
3) all expected counts ≥ 5 — 3) lowest exp count = 9 ≥ 5

conditions met --> χ^2 distribution --> χ^2 test for Homogeneity

Test Statistic:

$\chi^2 = \Sigma\frac{(obs - exp)^2}{exp} = \frac{(85 - 91)^2}{91} + \frac{(90 - 91)^2}{91} + \ldots =$

$= 8.547$

P-Value:

$P(\chi^2 > 8.547 | df = 4) = 0.0735$

Conclusion:

We fail to reject Ho b/c p-value of 0.0735 > alpha = 0.05. We have insufficient evidence that the distribution of defective products is not the same throughout the days of the week.

Worksheet in the notes, #2:

2) The following table is from the July 1993 publication of *Vital and Health Statistics* from the Centers for Disease Control and Prevention/National Center for Health Statistics. The individuals in the following table have only one of the three indicated irritations. Determine if the type of irritation is independent of the age group using a 0.05 level of significance.

Irritation	18-29	30-44	45-64	65+
Eye	440	567	349	59
Nose	924	1311	794	102
Throat	253	311	157	19

HW:

p. 643 #8, 10, 13, 14, 24

- Read each problem
- Determine which type of test (GOF, Homogeneity, Independence)
- Write the hypotheses for the test (Ho only)

Do problems #23, 29